AD-E501 763

Copy 24 of 35 copies

AD-A278 524

IDA DOCUMENT D-1474

A SURVEY OF FUZZY LOGIC AND APPLICATIONS

Irvin W. Kay

APR 19 1994

B D

January 1994

Approved for public release; distribution unlimited.

94-11671

5098

INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

94 4 18 098

IDA Log No. HQ 94-45020

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under IDA's Central Research Program. Its publication does not imply endorsement by the Department of Defense or any other Government Agency, nor should the contents be construed as reflecting the official position of any Government Agency.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY ***(Leave blank)***

2. REPORT DATE
January 1994

3. REPORT TYPE AND DATES COVERED
Finai--February–December 1993

4. TITLE AND SUBTITLE
A Survey of Fuzzy Logic and Applications

5. FUNDING NUMBERS
IDA Central Research Program

6. AUTHOR(S)
Irvin W. Kay

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311-1772

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Document D-1474

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 180 words)***

This document, prepared under IDA's Central Research Program, describes briefly some proposed or implemented applications of the theory of fuzzy sets. It also attempts to characterize the nature of the applications that have been successful and to suggest how the methodology might be more effectively applied to certain proposed analytical applications.

14. SUBJECT TERMS
fuzzy logic, fuzzy set theory, cluster analysis, signal processing

15. NUMBER OF PAGES
55

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

IDA DOCUMENT D-1474

A SURVEY OF FUZZY LOGIC AND APPLICATIONS

Irvin W. Kay

January 1994

INSTITUTE FOR DEFENSE ANALYSES
IDA Central Research Program

PREFACE

This document, prepared under IDA's Central Research Program, describes briefly some proposed or implemented applications of the theory of fuzzy sets. It also attempts to characterize the nature of the applications that have been successful and to suggest how the methodology might be more effectively applied to certain proposed analytical applications.

The material presented herein is published for the use and convenience of interested parties. It has not necessarily been completely evaluated and analyzed, nor subjected to formal IDA review.

CONTENTS

FIGURES

TABLES

I. INTRODUCTION

In 1965 L.A. Zadeh introduced the theory of fuzzy sets (Ref. 1) to deal with the kind of imprecise entities usually encountered in ordinary, nontechnical human discourse. His motivation was the possibility of formulating a design methodology for relatively inexpensive, nonoptimal, control systems that can perform complicated tasks.

With the advent of the general purpose microcomputer chip, a flood of applications of his methodology have appeared, ranging from smart household appliances for the consumer market to the automated control of industrial processing plants. In addition, applications of the theory to various kinds of analysis, such as medical diagnostics and signal processing, have been proposed.

This document briefly describes some of the applications, either proposed or already implemented, and outlines the basic concepts of the related fuzzy set theory. It also attempts to characterize the nature of the applications that have been successful and to suggest how the fuzzy methodology *might* be more effectively applied to certain proposed analytical applications.

II. APPLICATIONS

A. AUTOMATIC CONTROL SYSTEMS

Thus far, most of the commercially oriented applications of fuzzy logic appear to have been in the design of three types of automatic systems. In one the objective is to replace, at least partially, a human operator of a vehicle, such as an automobile, a railroad train, a seagoing vessel, or an aircraft. In the second it is to replace a human operator of a process plant, such as a cement kiln, a water purification processor, a sinter plant, or a box annealing furnace. In the third it is to make simple devices, such as household appliances, automatically perform certain changes during their operation that would ordinarily depend upon human judgment based on information acquired from experience, common sense, or intuition. A few of many examples that have been publicized are: elevator control (Fujitec/ Toshiba) based on an automatic evaluation of passenger traffic, selection of focal regions for a self-focusing camera (Canon/Nikon) or video camcorder (Sanyo Fisher/Canon), adjusting a vacuum cleaner's motor power (Matsushita) according to the floor condition and dust quantity, adjusting the heating element of a hot water heater (Matsushita) to the temperature and the amount of water being used.

In the first type of application the emphasis is on a dynamic control system that has one or more of the following characteristics: (1) its mathematical model is too complicated for the use of standard design methodology; (2) its operating environment is very noisy; (3) the available computer time and memory during its operation are both limited; (4) it must be inexpensive. In the second and third types the emphasis is on algorithms devised to implement nonquantitative human instructions, couched in subjective linguistic rather than objective numerical terms. In fact, these applications can also be classified as expert systems (or, in the case of the third type, perhaps mini-expert systems) the operation of which depends on rules derived from the advice of human experts like those that the systems are supposed to replace.

B. EXPERT SYSTEMS

As in the case of the process control plant examples mentioned earlier, fuzzy logic may be useful for transforming into rules that govern the output of more general types of

expert systems the kind of imprecise, subjective data acquired by humans from a combination of training and experience. For example, some analysts have suggested a role for fuzzy sets in designing computer driven expert systems to do, or at least assist in, medical diagnoses.

A few such systems, specialized to particular disorders, have been successfully implemented without the aid of fuzzy sets. However, they have taken years to develop, and a number of attempts to develop automatic diagnostic systems for other specific medical disorders have been unsuccessful. A later section (VI.C) will discuss a tool, enhanced by the use of fuzzy sets, that may be useful in dealing with certain difficult aspects of the general diagnosis problem.

C. PATTERN RECOGNITION

Among the announced applications of fuzzy logic is a handheld computer (Sony) that interprets handwritten input for data entry. This is an example of a pattern recognition problem that appears to be more amenable to fuzzy than to crisp algorithms.

The literature suggests that the fuzzy approach may be particularly advantageous in dealing with the initial, often difficult, requirement of feature selection, which is an important step in solving the pattern recognition problem. When the task is one that a human can perform, as in the case of handwriting interpretation, this is, in principle, an expert system application.

D. TARGET TRACKING

Suggestions, accompanied by a considerable amount of analysis and some testing by simulation, have been made in the literature for the application of fuzzy logic to automatic tracking of military targets. Because of its ability to take into account subjective information derived from human experience, in some circumstances the fuzzy approach should be useful in connection with the problem of track correlation.

E. SENSOR FUSION

Closely connected to the target tracking problem is a requirement, referred to in a more general context as sensor fusion, for processing, i.e., extracting relevant information from, the combined data contributed by a number of different sources. When this data includes subjective, qualitative, or estimated rather than measured, contributions from human observers, fuzzy methods seem appropriate.

F. SECURITY RISK ANALYSIS

An objective of security risk analysis is the assignment of a risk index to individual users of a secure system, such as a classified computer network. The purpose is to have a criterion for determining the type of access a user may have to certain parts of the network, e.g., particular communication channels, computer memory areas, or data files.

For simple systems the standard rules are easy to apply and require no computer assistance. However, for a very complex system that is host to several communities of interest and contains a variety of data bases covering an extensive classification range, it may be necessary to combine the effect of several rules to arrive at an appropriate risk index for each user. An example of such a system is the SDI National Test Bed.

Obviously, in any case the assignment of a risk index, which has an operational implication but no intrinsic meaning, is based to some extent on subjective judgment. Thus, the problem of determining the index in a complex environment by combining the effect of many factors, each of which has a subjective element, should be ideally suited to attack by the fuzzy logic approach. In fact, Mitre has proposed a method based on fuzzy logic for use in the risk analysis of the National Test Bed security system.

III. DEFINITIONS

A. FUZZY LOGIC

The fundamental entity of classical logic, dating back to Aristotle, is the proposition, which is defined to be any statement in ordinary discourse that is either true or false. This definition is sometimes referred to as "the law of the excluded middle" and the corresponding system of inference as a "two-valued logic," wherein every proposition has one of two possible truth values.

In the past logicians have proposed various multivalued generalizations. L. Zadeh's fuzzy logic concept (Ref. 1) is an example of such a generalization. It is a multivalued extension of the classical system to one that deals not only with propositions but also with statements in ordinary discourse that are meaningful, yet are neither true nor false in any precise sense.

Fuzzy logic deals primarily with subjective statements, such as "It was quite warm yesterday."[1] Although such statements do convey information, they are not true or false in any absolute sense; therefore, strictly speaking, they are not classical propositions. Nevertheless, it is possible to include them in logical discourse from which inferences can be drawn. Fuzzy logic practitioners have referred to such discourse as "common sense" argument.

B. FUZZY SETS

The inclusion of fuzzy propositions in logic leads to the idea of a fuzzy set, in which the requirements for membership are not precisely defined. An example might be the set of "HOT DAYS LAST NOVEMBER."

---

[1] Note, however, that the quasi-subjective statement "Yesterday was warmer than today" may have an absolute truth value, or it may not because of ambiguity in its meaning. Thus, the statement is definitely true if, in fact, the minimum temperature yesterday was greater than the maximum temperature today, or false if the maximum temperature yesterday was no greater than the minimum temperature today. Otherwise, unless made more precise by references to mean temperature, lowest temperature, etc., such a statement can have only a subjective implication.

By introducing the concept of partial membership, the fuzzy set definition accounts for imprecise requirements for the membership of an element in a set. This involves a generalization of the characteristic function of a classical set, for which membership is precisely defined.

The characteristic function of a classical set is a function having only two possible values: 0 or 1. Taking the value 1 at an element that is in the set and the value 0 at one that is not, it is defined over all elements in what is termed the "universe of discourse," which consists of all possible candidates for set membership.

The fuzzy set generalization is called the "membership function," and it can take on any value in the closed interval between 0 and 1. If the value is other than 0 or 1 at an element, the element is regarded as having partial membership in the set. To specify a fuzzy set completely, it is necessary to include its membership function in its definition.

A convenient notation for a fuzzy set A with membership function $\mu(u_i)$ is

$$S=\sum_{i=1}^{n}\mu_S(u_i)/u_i=\mu_S(u_1)/u_1+\mu_S(u_2)/u_2+\cdots+\mu_S(u_n)/u_n$$

when the universe of discourse is a finite fuzzy set U. If U is continuous the notation is

$$S=\int_U \mu_S(u)/u \quad .$$

The classical set, for which the membership function is the standard characteristic function that is either zero or one, is called a "crisp" set. It is obviously a special case of a fuzzy set.

It has been suggested that the membership function is nothing more than, or, at least, can be interpreted as, the probability that a given element belongs to the set. However, this interpretation obviously does not fit certain cases: e.g., the set of honest men. The assignment of a 0.75 membership in that set to someone does not necessarily imply an estimate of a 75 percent probability that he is honest. It could mean, for example, that he is known to be dishonest in his personal relationships but scrupulously honest in his business dealings, and it is estimated by those who know him best that he devotes three-quarters of his time to business and activities in which his integrity is irrelevant. More often, the index will be a personal judgment, amounting to a prediction that this would be the consensus response to a question starting with "On a scale from 0 percent to 100 percent, how would you rate the honesty of...?" in some hypothetical poll.

C. FUZZY VARIABLES

Variables, in the sense defined by Zadeh, are attributes that distinguish between elements of some universe of discourse. One example that he suggests is the color of an object, which belongs to a universe of objects.

Color defined by the wavelength of the radiation reflected from, or emitted by, an object is, in the usual sense of the term, a variable, each instance of which has a precise numerical value. However, in ordinary language color is not at all precise: as Zadeh observes, terms such as "red" or "blue" are labels for fuzzy sets of objects. They are fuzzy because for each the corresponding wavelength might take on any value in a range that is, itself, not well defined. Therefore, he regards color in the usage of ordinary language as a fuzzy variable.

The membership function of the fuzzy set corresponding to a value of a fuzzy variable defines the meaning of the fuzzy value explicitly. For example, for the universe of discourse consisting of numbers from 1 to 5, the values of a linear membership function μ_{SMALL} associated with the fuzzy set "SMALL" defined as the value of some fuzzy variable, e.g., lobster size, might be 1 if the size equals 1 lb, 0.75 if it equals 2 lb, 0.5 if it equals 3 lb, 0.25 if it equals 4 lb, and 0 if it equals 5 lb. Expressed as the membership function of a fuzzy set in the standard notation, the membership function of "SMALL" referring to lobster size could be written

$$\mu_{small} = 1/1+0.75/2+0.5/3+0.25/4+0/5$$

Of course, the membership function need not be linear. As in the case of the following example, a nonlinear function may represent human experience more appropriately.

In the usual applications of fuzzy methodology, membership functions defining the basic fuzzy sets that are involved, if not the result of expert knowledge, are easily derived from common experience. For example, a membership function for the set "HOT DAY" might reasonably be assigned the values given by the function

$$\mu(T) = \begin{cases} 1, & T > 85, \\ 1+\dfrac{(T-85)^3}{16000}, & T \geq 65, \\ \dfrac{(T-45)^3}{16000}, & T \leq 65, \\ 0, & T < 45. \end{cases}$$

where T is mean temperature measured in degrees Fahrenheit. The curve for the membership function according to this definition is shown in Figure 1.

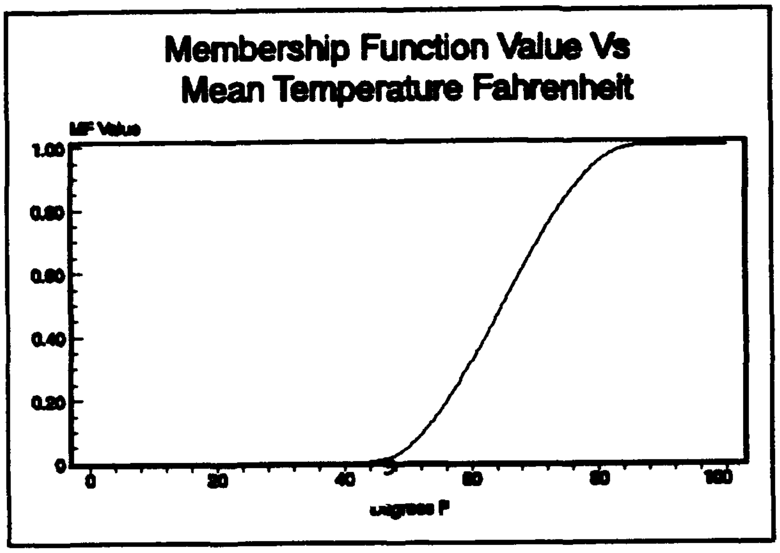

Figure 1. Membership Function for "HOT DAY"

D. LINGUISTIC VARIABLES

Zadeh distinguishes between a fuzzy variable, e.g., color, having values that are atomic terms, such as red and blue, and what he calls a linguistic variable, which is a type of fuzzy variable for which the values are language phrases. He gives height as an example of a linguistic variable if its values occur over a range of phrases, such as "not tall," "not very tall," "somewhat tall," "quite tall," "very tall," "very very tall."

E. FUZZY FUNCTIONS

Zadeh points out that the values of a fuzzy variable y can depend on those of a fuzzy variable x in a way that can be defined by simple conditional statements of the form if x is A then y is B; e.g., "if x is small then y is very large," "if x is not very small then y is very very large," etc.

F. FUZZY ALGORITHMS

Zadeh observes that the definition of fuzzy functional dependence by the use of simple conditional statements is analogous to the definition of ordinary functional dependence by the use of a table of associated function values, i.e., a table that assigns a

value of y for every value of x. On the other hand, another way to specify ordinary functional relationships is to use algorithms, e.g., in the form of an algebraic formula or, more generally, a sequence of computational steps.

Similarly, instead of using a list of conditional statements to specify a fuzzy functional dependence, it is often possible to accomplish the purpose by means of a fuzzy algorithm, which can also be used to define other fuzzy constructs such as sets. According to Zadeh's definition, a fuzzy algorithm is an ordered sequence of instructions, some of which may contain fuzzy set labels. That is, a fuzzy algorithm can have a se[illegible]nce of instructions such as "if x is small decrease y a little," "if x is large increase y t[illegible] when x is 10 and y is less than 10 then change y to 10."

IV. UTILITY OF FUZZY LOGIC

A. WHEN THE FUZZY APPROACH IS USEFUL

Fuzzy logic may be useful in the following circumstances: (1) when a human observer supplies data that is subjective or is couched in imprecise language; (2) when mathematically precise algorithms are too complicated or otherwise inappropriate; (3) when it is desirable to model ordinary human behavior; (4) when it is necessary to deal with concepts that cannot be measured but must be compared in a logically consistent way. The vast literature concerned with fuzzy methodology now includes many articles describing in detail examples of such applications, a considerable number of which have been patented.

According to some critics, classical probability theory can accomplish whatever fuzzy logic can. This belief, which does not appear to be justified by either theoretical considerations or existing applications of the fuzzy approach, seems to be based on the idea that the same results can always be achieved with Bayesian decision theory.

However, in many fuzzy logic applications (e.g., fuzzy control systems, signal processing) decision theory plays no role. In addition, the Bayesian methodology has a well-known weakness: the need for what is often an arbitrary assignment of a priori probabilities to certain events. It is usually assumed that these probability assignments are expressions of human experience and therefore, to some extent, represent real data inputs; but in many instances what are regarded as data amount to no more than guesses.

On the other hand, decision methodology based on fuzzy logic does not depend on substituting guesses for real data. It should deal in a well-organized, logically consistent way with information derived from human experience, but only when such information actually exists.

B. HOW FUZZY LOGIC IS USED

Perhaps the most important advantage of fuzzy methodology is the existence of well-defined mathematical set operations analogous to the standard operations of Boolean algebra used with ordinary crisp sets. The fuzzy set operations satisfy most of the same basic axioms, as well as DeMorgan's theorem. Starting with a few simple, easily

specified, basic fuzzy sets it is therefore possible to generate complex fuzzy sets in a logically consistent way and by this means to formulate nontrivial, well conceived algorithms that, nevertheless, make use of subjective or otherwise imprecise information.

The following definitions of fuzzy set operations are generally recognized in the literature. It should be kept in mind that the membership function of a fuzzy set determines the fuzzy set, just as the characteristic function of an ordinary crisp set determines the crisp set.

1. Union

The *union* of two fuzzy sets A and B with membership functions $\mu_A(m)$ and $\mu_B(m)$ is given by the membership function defined for each value of m as the larger of $\mu_A(m)$ and $\mu_B(m)$, i.e., by

$$\mu_{A \cup B}(m) = \max\,[\mu_A(m), \mu_B(m)] \quad .$$

For example, suppose, as in a previous example, the fuzzy set "SMALL" is defined in terms of a membership function μ_{SMALL}, over the universe of discourse consisting of the integers 1, 2, 3, 4, 5, by Table 1.

Table 1. Membership Function for the Fuzzy Set "SMALL"

SMALL	1	2	3	4	5
μ	1.00	0.75	0.50	0.25	0.00

and the fuzzy set "MEDIUM" is defined, in a similar manner by Table 2.

Table 2. Membership Function for the Fuzzy Set "MEDIUM"

MEDIUM	1	2	3	4	5
μ	0.50	0.75	1.00	0.75	0.50

Then the *union* set "MEDIUM *or* SMALL" is given by Table 3.

Table 3. Membership Function for the Fuzzy Set "MEDIUM *or* SMALL"

MEDIUM *or* SMALL	1	2	3	4	5
μ	1.00	0.75	1.00	0.75	0.50

2. Intersection

The *intersection* of two fuzzy sets A and B with membership functions $\mu_A(m)$ and $\mu_B(m)$ is given by the membership function defined for each value of m as the smaller of $\mu_A(m)$ and $\mu_B(m)$, i.e.,

$$\mu_{A \cap B}(m) = \min[\mu_A(m), \mu_B(m)] \quad .$$

If the sets A and B in this case are the sets "SMALL" and "MEDIUM" used in the previous example, illustrating the *union* operation and given by Tables 1 and 2, then the set "MEDIUM *and* SMALL," which may be understood as equivalent to the *intersection* of "MEDIUM" and "SMALL," will be given by Table 4.

Table 4. Membership Function for the Fuzzy Set "MEDIUM *and* SMALL"

MEDIUM *and* SMALL	1	2	3	4	5
μ	0.50	0.75	0.50	0.25	0.00

3. Complement

Relative to a specific universe of discourse, the *complement* of a fuzzy set A, written $\overline{A}$, has a membership function that, because it must reduce to the classical definition in the case of a crisp set with a characteristic function for a membership function, is given by

$$\mu_{\overline{A}}(m) = 1 - \mu_A(m) \quad ,$$

where $\mu_A(m)$ is the membership function of A. For example, if A is the previously defined fuzzy set "SMALL," for which the membership function is given by Table 1, then $\overline{A}$ is the fuzzy set "NOT SMALL" and the membership function of $\overline{A}$ is given by Table 5.

Table 5. Membership Function for the Fuzzy Set "NOT SMALL"

NOT SMALL	1	2	3	4	5
μ	0.00	0.25	0.50	0.75	1.00

Note that the Boolean operations *intersection*, *union*, and *complement* applied to fuzzy sets in some cases lead to results that are quite different from what we would expect if the sets were crisp. For example, (1) the membership function for the *intersection* of the fuzzy sets "SMALL" and "NOT SMALL" is 0/1 + .25/2 + .50/3 + .25/4 + 0/5, which, because it indicates that some membership in the set is possible, violates the law of the excluded middle of classical two-valued logic; (2) the membership function given in

Table 5 for the set "NOT SMALL" also provides a reasonable definition of "LARGE" rather than "MEDIUM *or* LARGE," which would be implied if the sets "SMALL," "MEDIUM," "LARGE" were crisp and therefore mutually exclusive.[2]

4. Linguistic Hedges

In addition to the Boolean operations, linguistic hedges can also be used to generate new fuzzy sets. These are modifiers, such as "very" and "rather," that alter basic sets, such as "SMALL" and "LARGE."

The commonly accepted operational meaning of "very" is that it alters a membership function by squaring each of the values that the function takes over the universe of discourse. Thus, according to Table 2 the membership function of "VERY SMALL" should be given by Table 6.

Table 6. Membership Function for the Fuzzy Set "VERY SMALL"

VERY SMALL	1	2	3	4	5
μ	1.00	0.56	0.25	0.06	0.00

The hedge "rather" changes a membership function by translating it along the universe of discourse by an appropriate amount r. Thus, to obtain the membership function for "RATHER SMALL" with "SMALL" given by Table 1, an appropriate value for r might be 1. Then the membership function for "RATHER SMALL" would be given by Table 7.

Table 7. Membership Function for the Fuzzy Set "RATHER SMALL"

RATHER SMALL	1	2	3	4	5
μ	1.00	1.00	0.75	0.50	0.25

2 If the sets were crisp the *complement* of "SMALL" would be "MEDIUM" *or* "LARGE," which, for any definition of "LARGE" would have a different membership function than the one given in Table 5. That is, the *union* of "MEDIUM" and "LARGE" would (according to Table 2) assign membership values of at least .50 to 1, .75 to 2, and 1 to 3. In this case the most reasonable membership function for "LARGE" would be .50/1 + .75/2 + 1/3 + 1/4 + 1/5, which seems less satisfactory than assigning the values in Table 5 to the membership function of "LARGE."

5. Fuzzy Conditional Statements

In control systems both the fuzzy relations between states and the fuzzy rules defining a control algorithm or function have the form of a conditional statement[3] "if A then B." If the conditional statement is interpreted as a relation R between fuzzy sets A and B, defined in different universes of discourse, then R can be represented as a kind of Cartesian product, symbolically:

$$R = A \times B \ .$$

This Cartesian product is a fuzzy set with a membership function that depends on two independent variables, each of which ranges over a different universe of discourse.

The membership function of R is actually formed by combining all possible pairs of values from the membership functions of A and B as if they were due to an *intersection* of two sets.[4] That is, it is given by

$$\mu_R(m_1, m_2) = \mu_{A\times B}(m_1, m_2) = \mu_{A\cap B}(m_1, m_2) = \min[\mu_A(m_1), \mu_B(m_2)] \ .$$

It is convenient to express the membership function in matrix form, subsequently treating the matrix as a representation of R:

$$R = A\times B = \begin{bmatrix} \min[\mu_A(m_{11}), \mu_B(m_{21})] \ldots \min[\mu_A(m_{11}), \mu_B(m_{2r})] \\ \vdots \qquad\qquad\qquad \vdots \\ \min[\mu_A(m_{1s}), \mu_B(m_{21})] \ldots \min[\mu_A(m_{1s}), \mu_B(m_{2r})] \end{bmatrix} .$$

As an example of how the matrix elements are calculated in a specific case, consider the fuzzy conditional statement "If X is small then Y is medium," where Tables 1 and 2 give the membership functions of "SMALL" and "MEDIUM." Substituting the values from those tables into the matrix representation for R, replacing each element by the designated minimum, leads to

$$R = \begin{bmatrix} 0.50 & 0.75 & 1.00 & 0.75 & 0.50 \\ 0.50 & 0.75 & 0.75 & 0.75 & 0.50 \\ 0.50 & 0.50 & 0.50 & 0.50 & 0.50 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

---

3 The conditional statements need not be simple: A and B may be complex fuzzy sets formed by the use of Boolean operations to combine and otherwise alter a number of simple sets.

4 If the sets were crisp the Cartesian product would, in fact, be the *intersection* of two "cylindrical" sets, and the associated characteristic function would depend on two arguments.

If the sets A = "SMALL X" and B = "MEDIUM Y" were crisp instead of fuzzy the matrix elements would consist of ones and zeroes, derived from the characteristic functions of the cylindrical extensions of A and B to the two-dimensional universe of discourse in which both can participate. The distribution of matrix elements would represent the characteristic function of the *intersection* of the cylindrical extensions of A and B; i.e., the ones would represent the elements that the two cylindrical sets have in common.

Although the modifiers "small" and "medium" refer to universes of discourse having the same number of values, i.e., five, the two universes of discourse are completely different. The fact that there are five possible values of X and five possible values of Y should be regarded as a coincidence. If the numbers were different R would not be a square matrix, but the computation in the example would otherwise be carried out in exactly the same way.

In applying the rule "if A then B" in a control algorithm the observed value of X is assumed to be fuzzy and therefore must be assigned an appropriate membership function based on the designer's knowledge or intuition about the relative precision of the observation. The contribution of the rule will then depend on the extent to which the fuzzy set corresponding to the observed value of X coincides with the set A, i.e., on the intersection of the two sets, for each value of the membership function for the set B in the relation R.

In terms of the matrix representing R, the elements in every column are the membership function values of A associated with some value of the membership function of B. Thus, if A_0 is the fuzzy set corresponding to the observed value of X, the outcome of the rule will depend on the intersection of A_0 with the sets whose membership function values are the elements in each column of the matrix. Logically, the union of these intersections should determine the consequence B_0 of the rule, given the observation A_0.

In fact, the procedure called *composition* normally used to determine the set B_0 that the rule implies (cf. Ref. 2 for a detailed example) consists of the following steps: (1) calculate the membership functions corresponding to the *intersection* of A_0 with each set for which a column of the matrix representing R is the associated membership function; (2) calculate the membership function associated with the *union* of the sets corresponding to the membership functions calculated in step (1).

Symbolically, the composition procedure can be written

$$B_0 = A_0 \circ R \rightarrow \mu_{B_0}(m_2) = \max_{m_1} \min[\mu_{A_0}(m_1), \mu_R(m_1, m_2)] \quad .$$

It is evident from this expression that composition leads to the result $B_0 = B$ if $A_0 \cap A = A$, which is logically consistent.

As an example, suppose that the conditional statement "If the amount of material is small then a medium amount of heat should be applied" is a rule furnished by a human expert as part of a fuzzy algorithm designed to automate some process. For convenience, assume that the previously defined membership functions for "SMALL X" and "MEDIUM Y" and the matrix representation of the associated relation R are applicable in this example.

Suppose at some point during the process the amount of material characterized by the rule statement as small actually appears to be more closely approximated by the membership function (corresponding to A_0) specified in Table 8 rather than the one given in Table 1.

Table 8. Membership Function for a Specified Amount of Material

SPECIFIED MATERIAL	1	2	3	4	5
μ	0.70	1.00	0.90	0.50	0.25

Then

$$\min[\mu_{A_0}(m_1), \mu_R(m_1, m_2)]$$

$$= \begin{bmatrix} \min[.70,.50] & \min[.70,.75] & \min[.70,.10] & \min[.70,.75] & \min[.70,.50] \\ \min[1.0,.50] & \min[1.0,.75] & \min[1.0,.75] & \min[1.0,.75] & \min[1.0,.50] \\ \min[.90,.50] & \min[.90,.5] & \min[.90,.50] & \min[.90,.50] & \min[.90,.50] \\ \min[.50,.25] & \min[.50,.25] & \min[.50,.25] & \min[.50,.25] & \min[.50,.25] \\ \min[.25,0.0] & \min[.25,0.0] & \min[.25,0.0] & \min[.25,0.0] & \min[.25,0.0] \end{bmatrix}$$

$$= \begin{bmatrix} .50 & .70 & .70 & .70 & .50 \\ .50 & .75 & .75 & .75 & .50 \\ .50 & .50 & .50 & .50 & .50 \\ .25 & .25 & .25 & .25 & .25 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

completes step (1) of the calculation. Step (2) requires picking the maximum element in each of the columns of the last matrix, leading to the result

$$\mu_{B_0} = [.50 \quad .75 \quad .75 \quad .75. \quad 50]$$

for the membership function of the set "APPLIED HEAT" that corresponds to the membership function given by Table 8 for the set "SPECIFIED MATERIAL."

A fuzzy control algorithm generally consists of a sequence of conditional statements: "If A_1 then B_1," ..., "If A_n then B_n." It can be regarded as equivalent to a single instruction formed by inserting the term "else" (meaning *or*) in front of every statement in the sequence but the first. Thus, the algorithm can be interpreted as the logical *union* of its constituent conditional statements.

It follows that the relation matrix R for the entire algorithm can be calculated in a simple manner in terms of the relation matrices R^i associated with the separate conditional statements. Each element of R must be the maximum value of all corresponding elements of the R^i, i.e.,

$$R_{jk} = \max_i [R^i_{jk}] \quad .$$

As an example, suppose that a fuzzy control algorithm consists of the two conditional statements: (1) if the amount of material is small then a medium amount of heat should be applied; (2) if the amount of material is large then a large amount of heat should be applied. The relation R^1 implied by statement (1) might be represented by the matrix previously calculated in connection with that statement. The matrix representing the relation R^2 implied by statement (2) might be constructed in the same way from the entries in Table 5 interpreted as referring not only to a large amount of material but also to a large amount of heat,[5] and the result is

$$R^2 = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.00 & 0.25 & 0.50 & 0.50 & 0.50 \\ 0.00 & 0.25 & 0.50 & 0.75 & 0.75 \\ 0.00 & 0.25 & 0.50 & 0.75 & 1.00 \end{bmatrix} \cdot$$

Thus,

$$R = R^1 \cup R^2 = \begin{bmatrix} 0.50 & 0.75 & 1.00 & 0.75 & 0.50 \\ 0.50 & 0.75 & 0.75 & 0.75 & 0.50 \\ 0.50 & 0.50 & 0.50 & 0.50 & 0.50 \\ 0.50 & 0.25 & 0.50 & 0.75 & 0.75 \\ 0.50 & 0.25 & 0.50 & 0.75 & 1.00 \end{bmatrix} \cdot$$

5 The membership function for "LARGE AMOUNT OF MATERIAL" could be completely different from the membership function for "LARGE AMOUNT OF HEAT." In the example given here, the two functions are assumed to be identical for simplicity. Of course, the independent variables in the membership functions are inherently different since the measurement units pertaining to an amount of material are, in any case, different from those pertaining to an amount of heat.

To implement the control function the fuzzy algorithm must determine the amount of heat to be applied when the amount of material is a given value. Suppose that the amount of material used in a particular instance of the process is 2. Then the second row of the matrix representing R is the membership function of the fuzzy control set specifying the amount of heat to be applied. The values of the function are displayed in Table 9.

Table 9. Membership Function for the Fuzzy Amount of Heat to be Applied

AMOUNT OF HEAT	1	2	3	4	5
μ	0.50	0.75	0.75	0.75	0.50

Of course, to actually implement the control function it is necessary to derive a single, non-fuzzy, value for the amount of heat. Reference 2 suggests two possible ways to replace a fuzzy control set with a single number: (1) choose the amount corresponding to the maximum membership value, averaging over the corresponding amounts if several maxima exist; (2) take a weighted average of the amounts, wherein the weights are the membership function values. The result obtained from method (1) is clearly 3. The result obtained from method (2),

$$\frac{\sum_{n=1}^{5} n\mu(n)}{\sum_{n=1}^{5} \mu(n)} = \frac{.5+2\times.75+3\times.75+4\times.75+5\times.5}{.5+.75+.75+.75+.5} = 3 \ ,$$

is the same as that obtained from (1); therefore, in this case it makes no difference which of the two methods is used. Although no clear reason exists to choose one method over the other, (2) seems to be the preferred one in the literature.

A variation of the fuzzy control function is the predictive fuzzy control, used in Ref. 3 for the design of an automatic train operating system. In this approach the rules, which are still based on conditional statements, have the form: "If 'A is X' leads to 'B is Y' then 'C is Z'." In other words, the outcome of a fuzzy prediction determines the applicable fuzzy rule.

V. SIGNAL PROCESSING

Reference 4 gives an example of a fuzzy logic application to signal processing. The authors refer to the specific method presented in their paper as Heuristically Constrained Estimation (HCE). They characterize HCE as a way to improve the effectiveness of a standard signal processing operation by introducing constraints based on subjective information derived from experience and common sense reasoning.

Because the language defining them is imprecise, such constraints are fuzzy; i.e., they are expressible in terms of a membership function $\mu_C(x)$ of a fuzzy set. Reference 4 suggests that the subjective information given by this membership function can be incorporated into a standard signal processing technique based on maximizing some objective function j(x)[6] by multiplying j(x) by the function $\mu_C(x)$ and using the result in place of j(x).

If the constraint were crisp rather than fuzzy, $\mu_C(x)$ would reduce to the characteristic function, defined to have the value zero when x is forbidden by the constraint and the value one when x is allowed. Then multiplying j(x) by it would have the effect of enforcing the constraint because it would rule out x as a possible maximizing element whenever the constraint eliminated x, while otherwise it would leave the value of j(x) unchanged. Since the fuzzy set membership function is a generalization of the crisp set characteristic function, multiplying j(x) by $\mu_C(x)$ is a natural way to generalize from a crisp to a fuzzy constraint.

For a simple example to illustrate how a fuzzy constraint might improve the result in a typical parameter estimation problem, suppose that the objective of a ground-based radar is to estimate the range and radial velocity of a target by processing the signal returned from it. Suppose, also, that because of terrain characteristics at the radar location, it is known that only targets located beyond a certain range can possibly be moving at high velocities. This knowledge can be used to introduce a fuzzy constraint, which should improve range and velocity estimates for some targets.

---

[6] For example, in the case of maximum likelihood estimation j(x) would be the conditional probability $p(z|x)$ that the observed data is z, given that the correct data is x.

The time delay τ of the radar return signal determines the range estimate R by

$$R=\frac{c\tau}{2},$$

where c is the speed of light. The Doppler frequency shift Δf determines the velocity estimate V by

$$\Delta f=\frac{2V}{c}f,$$

where f is the carrier frequency of the radar signal.

The standard method of estimating τ and Δf is to locate the maximum of the so-called ambiguity function (Ref. 5) given by

$$\theta(\tau',\omega')=\left|\int_{-\infty}^{\infty}\tilde{f}(z-\frac{\tau'}{2})\,\tilde{f}^*(z+\frac{\tau'}{2})\,e^{i\omega' z}dz\right|^2,$$

where $\tilde{f}$ is the waveform normalized so that

$$\int_{-\infty}^{\infty}|\tilde{f}(z)|^2dz=1,$$

$\tilde{f}^*$ is its complex conjugate, and, for τ_a representing the actual value of τ and $\omega_a=2\pi\Delta f_a$ the actual value of the Doppler shift,

$$\tau'=\tau-\tau_a,\ \ \omega'=\omega-\omega_a.$$

Because the waveform is normalized

$$\theta(0,0)=1,$$

whereupon it follows from Schwartz's inequality[7] that the ambiguity function has its maximum value, which is 1, when the Doppler frequency and range variables are both equal to their actual values.

Usually the ambiguity function is due to a sequence of many pulses, causing it to have a number of local maxima periodically distributed over a large region of the range-Doppler plane. These maxima are generally smaller than the maximum associated with the point whose coordinates are equal to the actual range and Doppler values. However, additive noise in the received signal can sometimes increase the amplitude of the ambiguity

---

7 I.e., . $$\left|\int_{-\infty}^{\infty}F(x)\,G(x)\,dx\right|^2\le\int_{-\infty}^{\infty}|F(x)|^2dx\int_{-\infty}^{\infty}|G(y)|^2dy.$$

function at one or more of these additional local maxima enough to make it exceed the amplitude corresponding to the correct range and Doppler values.

It may be possible to avoid the resulting error by making use of any available information to restrict the region where the actual range and Doppler values can be located. Sometimes the available information for this purpose may be fuzzy rather than precise, as in the case of the present example. The suggested procedure described in what follows is one way to incorporate such information into the signal processing.

The knowledge that if a target velocity is large its range must be large is the assumed fuzzy constraint. The membership function corresponding to the constraint will depend on the two variables: τ and Δf.

The associated fuzzy constraint set consists of τ and Δf values implied by the condition "if Δf is large then τ is large." This condition defines the set

$$S_{const}(\tau,\Delta f) = [(large\,\Delta f)\cap(large\,\tau)]\cup\overline{(large\,\Delta f)}.$$

To find the constraint membership function, the first step is to define membership functions for the sets "LARGE τ" and "LARGE Δf." The following definitions may be appropriate:

$$\mu_{Large\,\tau}(\tau) = \begin{cases} 0, & R<10, \\ 1-\dfrac{b}{\tau^2}, & R\geq 10, \end{cases}$$

where τ is in seconds, the range R is in nautical miles, and

$$b = 1.528\times10^{-8}\,sec^2;$$

$$\mu_{Large\,\Delta f}(\Delta f) = \begin{cases} 0, & V<100, \\ 1-\dfrac{a}{\Delta f^3}, & V\geq 100, \end{cases}$$

where Δf is in Ghz, the carrier frequency f is 10 Ghz, the radial velocity V is in knots, and

$$a = 4.05\times10^{-17}\,Ghz^3.$$

Figures 2 and 3 present curves depicting these membership functions, but in terms of range and radial velocity rather than τ and Δf.

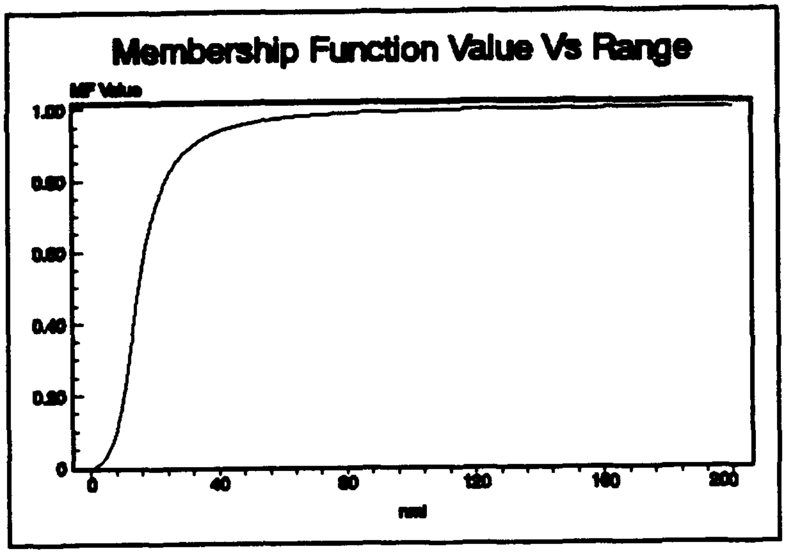

Figure 2. Membership Function for "LARGE RANGE"

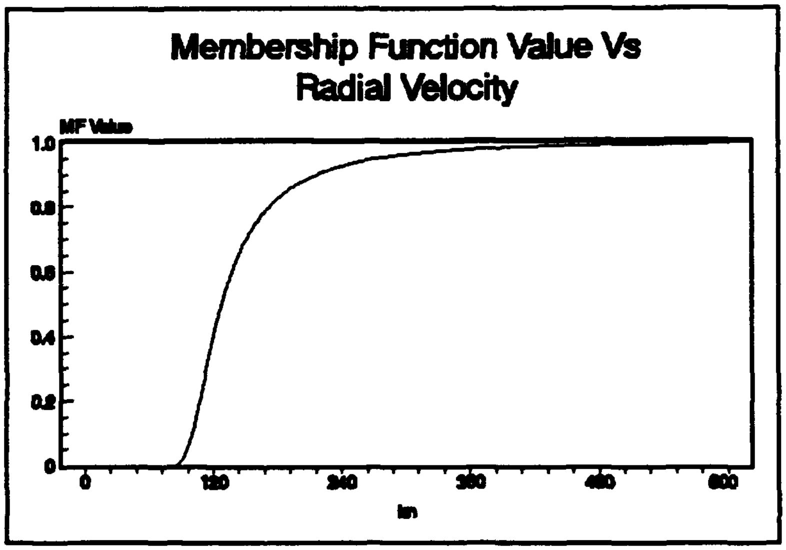

Figure 3. Membership Function for "LARGE VELOCITY"

The membership function for the constraint set $S_{const}(\tau, \Delta f)$ can be calculated in terms of the membership functions for the "LARGE τ" and "LARGE Δf" sets by means of the fuzzy rules cited earlier. Thus, defining

$$\mu_1(\tau, \Delta f) = \min[\mu_{large\tau}(\tau), \mu_{large\Delta f}(\Delta f)]$$

and

$$\mu_2(\Delta f) = 1 - \mu_{large\Delta f}(\Delta f),$$

those rules lead to the following expression for the constraint set membership function:

$$\mu_{const}(\tau, \Delta f) = \max[\mu_1(\tau, \Delta f), \mu_2(\Delta f)],$$

corresponding to $S_{const}(\tau, \Delta f)$. All that is needed to take into account the fuzzy constraint is to replace the original ambiguity function $\theta(\tau,\omega)$ with a new ambiguity function $\theta'(\tau,\omega)$ given by

$$\theta'(\tau, \omega) = \theta(\tau, \omega)\, \mu_{const}\left(\tau, \frac{\Delta f}{2\pi}\right).$$

The particular application that Ref. 4 addresses is the enhancement of an important tool used in oil exploration: the detection of layered geological structures in the earth by means of seismic waves. The problem is to estimate the distances between the interfaces separating the layers of different materials that a signal encounters as it propagates, as well as the change in the acoustic impedance across each interface.

An explosive source generates the seismic signal, which evolves into a convolution of the initial waveform with an impulse response wave train consisting of a series of sharp spikes caused by reflections from the layer interfaces and referred to as events. The separation between two successive spikes will be roughly proportional to the length of the propagation path that the signal follows between the interfaces bounding a layer. The amplitudes of the spikes will be monotonically related to the impedance changes at the layer interfaces. When its amplitude is large a spike is called a "significant event."

The processing task is to deconvolve the signal, which is corrupted by the addition of a noise term. Given the statistical behavior of the noise, it is possible to derive an optimum filter or algorithm for this purpose.

However, by taking into account additional, independent, knowledge about the interaction of the signal with its propagation environment, it is also possible to improve the result that the optimum algorithm would otherwise produce. In fact, Ref. 4 demonstrates

that the following common sense observations concerning the events that can occur in the deconvolved signal offer such knowledge.

1. The number of significant events contributing to the observed signal will be relatively small.
2. Too many events should not occur in any subsequence of length L.
3. Significant events should not occur close together in any subsequence of length L.

The common sense argument for observation 1 is that, to be useful, the geological description of the earth should be given in terms of a small number of distinct layers. Observation 2 is equivalent to the statement that many layers should not be packed in a small region, corresponding to the subsequence of length L. Observation 3 is equivalent to the statement that adjacent layers with greatly differing impedances should not be very thin.

Two quantities are to be chosen subjectively: (1) the number of significant layers to consider as appropriately descriptive of the geological earth model, and (2) the length L of the subsequence of events to be considered as corresponding to the time between reflections from the interfaces bounding a relatively thin layer. That is, experience determines the choice of these two parameters. Alternatively, the first might be implemented as a fuzzy variable. The common sense observations, themselves, constitute fuzzy constraints in the signal processing application.

It is useful to regard an event as the detection of a signal amplitude above a certain threshold within a small specified time interval. In terms of such events Ref. 4 defines membership functions for some elementary fuzzy sets corresponding to concepts *close*, *significant*, and *too many*, which were introduced in defining the fuzzy constraints.

For this purpose the first step is to prescribe the domain of discourse associated with each concept. The domain of discourse for the first concept consists of all pairs of events, indexed by integers i and j, corresponding to pairs of spikes in a wave train. For the second it consists of events in all possible subsequences s of length L. For the third it consists of the subsequences s. The remaining step is to define the associated membership functions over these domains of discourse:

$$\mu_{close\text{-}pair}(i,j) = \max\left[\frac{15-|i-j|}{15}, 0\right],$$

$$\mu_{significant\text{-}event}(event \in s) = \frac{|amplitude\ of\ event \in s|}{|amplitude\ of\ max\ event \in s|},$$

$$\mu_{too\text{-}many}(s) = \frac{1}{L}\sum_{i=1}^{L} \mu_{significant\text{-}event}(event\ i)\ .$$

Reference 4 then uses fuzzy Boolean operations on these membership functions to define membership functions associated with more complex concepts. Thus,

$$\mu_{significant\text{-}pair}(i,j) = \mu_{significant\text{-}event(i) \cap significant\text{-}event(j)}(i,j)$$
$$= \min[\mu_{significant\text{-}event}(i), \mu_{significant\text{-}event}(j)]\ .$$

Also, defining a subsequence of close significant events by

$$cl\text{-}sig\text{-}ev(s) = \bigcup_{i,j \in s} [sig\text{-}pair(i,j) \cap cl\text{-}pair(i,j)]\ ,$$

it follows that

$$\mu_{cl\text{-}sig\text{-}ev}(s) = \max_{ev\ pairs\ i,j \in s} [\min(\mu_{sig\text{-}pair}(i,j), \mu_{cl\text{-}pair}(i,j))]\ .$$

Reference 4 is then able to define a fuzzy constraint set C in terms of the set B of all subsequences s of any sequence x of events:

$$C = \bigcap_{s \in B} \overline{too\text{-}many(s)} \cap \overline{cl\text{-}sig\text{-}ev(s)}\ .$$

With the aid of DeMorgan's theorem, i.e., the identity

$$\overline{A \cup B} = \overline{A} \cap \overline{B},$$

it can be seen that the fuzzy constraint is equivalent to a membership function given by

$$\mu_C(x) = \min_{s \in B} [1 - \max[\mu_{too\text{-}many}(s), \mu_{cl\text{-}sig\text{-}ev}(s)]]$$

Because of its complexity, Ref. 4 does not describe in detail the algorithm used to carry out the optimization of the product $j(x)\mu_C(x)$. However, standard numerical analysis algorithms exist for calculating functions that maximize such functionals.

The method of incorporating fuzzy constraints in the two foregoing examples does not appear to be applicable to least mean square error filtering, e.g., to the design of Kalman or Wiener filters. Nevertheless, it should be possible to enhance the performance of this kind of signal processing by introducing fuzzy constraints in some fashion.

A common way to model a Kalman or Wiener filter is to regard it as a dynamic control system, defined by a system of differential equations for a set of state variables. Therefore, using the flow diagram representation of the process as a guide, it may be possible to introduce fuzzy constraints into a least square error filter design in the same way that fuzzy controls have been introduced in numerous real control system applications, cf. Ref. 6.

VI. FUZZY CLUSTER ANALYSIS

A. NON-FUZZY CLUSTERS

Cluster analysis is a method of organizing data by classifying certain data entities called "subjects" into sets called "clusters" according to features, called "attributes," that the subjects have in common. In many cases the attributes are defined in terms of a number of characteristic entities, called "targets," associated with the subjects. In these cases the term "affinities" may be used instead of "attributes."

For example, one kind of zoological taxonomy classifies mammals according to the number of teeth that an animal may have in each of eight possible dental categories. These consist of the upper and lower of each of four categories: incisors, canines, premolars, and molars. In this case the subjects are the mammal types, the targets are the eight tooth classifications, and the number of teeth that a mammal has in each are the affinities or attributes.

Another example is the use of cluster analysis to determine the potential for the existence of political blocks, using data from interviews in which spokespersons for various organizations or institutions are asked what positions their organizations will take on a list of controversial issues. The subjects in this case are the organizations, the targets are the issues, and the affinities are the positions that the organizations take on the issues.

In these examples the number of clusters would be unknown beforehand. In other cases the number of clusters might be prescribed rather than determined by analysis.

For example, in a cluster analysis applied to medical data the purpose may be to discover how to distinguish between some number of specified diseases by means of a specified set of symptoms. The data would consist of the symptoms experienced by every patient in a group of patients, each of whom would already have been diagnosed by means of laboratory tests as having a particular disease. In this example the patients are the subjects, the symptoms are the targets, and the affinities of a subject, i.e., patient, are the data on which symptoms the patient does or does not have. Every cluster is predefined to be the set of patients having one of the specified diseases, and the objective of the analysis

is equivalent to determining what set of affinities, i.e., symptoms and lack thereof, also defines each of the clusters.

B. FUZZY CLUSTERS

In standard cluster analysis the clusters are well defined sets. However, in many circumstances, perhaps more often than not, membership in a cluster may be ambiguous. This is obviously the case if the defining affinities are fuzzy, but it may also be useful to assume that the clusters are fuzzy sets even when the affinities are crisp. A typical example is the classification of mammals by the number of their teeth in each of eight dental categories.

In that example, as in most others, the number of clusters is large and the sizes of the separate clusters are small if the requirement for membership is that the subjects have a high percentage of the possible affinities, e.g., seven out of eight, in common. Relaxing the requirement will reduce the number of clusters and increase the membership of some of them.

But with no relevant information aside from the original data, the choice of a specific number of common affinities for the membership requirement would be somewhat arbitrary. Therefore, regarding sequences of possible clusters as one fuzzy set would, in effect, make use of all of the available data while ensuring that the analysis remained objective.

One method of accomplishing this is to regard each separate cluster as generated by a particular subject, identified as a "linch pin."[8] That is, membership in the cluster will depend on subjects having the same affinities as the linch pin. The most efficacious linch pins to associate with the different clusters are those having the largest number of affinities in common with other subjects, but not with each other. A table of values between zero and one and proportional to the number of affinities each subject has in common with the linch pin of a fuzzy cluster provides a reasonable definition of the membership function that defines the cluster.

---

8 The so-called leader algorithm (Ref. 8) for standard, non-fuzzy clustering, is based on a similar point of view. Assuming some definition of distance between subjects, it builds each cluster around a particular subject, called a leader, by assigning other subjects to the cluster if their distance from the leader is sufficiently small. Obviously, the terms "leader," used in Ref. 7, and "linchpin," used here and in certain social science contexts, refer to identical concepts. Reference 7 also uses the term "case" instead of "subject."

A simple algorithm for choosing the linch pins starts by constructing the square, symmetric matrix with elements A_{ij} equal to the number of affinities that subject i has in common with subject j, where both i and j enumerate all of the subjects. The first linch pin, subject l, chosen by the algorithm is such that, for some k different from l,

$$A_{lk} = \max_{i \neq j} A_{ij} \ .$$

Row l of the matrix determines the membership function defining the corresponding cluster. Since a subject has at most one affinity for every target, a membership function value is the ratio of the number of affinities that a subject has in common with the linch pin to the total number of targets. The algorithm chooses the linch pins that determine successive clusters in the same way as it did the first, but in each case the choice is only from among those subjects with low membership values in the previously defined clusters.

Obviously the algorithm fosters some ambiguity in the definition of clusters since the specification of what constitutes a low enough membership in a previously determined cluster to permit a subject to be a linch pin candidate is somewhat arbitrary. Also, the selection of linch pins, as well as the clusters that they determine, will sometimes depend upon the order in which the subjects happen to appear as elements in the subject matrix.

However, in neither case is the ambiguity completely arbitrary. In the first case it can be regarded as another fuzzy variable, to be interpreted accordingly. In the second case its occurrence is usually evidence of competing alternatives actually existing in the real world phenomena being investigated, implying that further research with this possibility in mind may be worthwhile.

C. EXAMPLES

1. Medical Diagnosis

Reference 7 discusses a potential application of fuzzy sets in cluster analysis to an automatic medical diagnosis of a stomach disorder from a list of symptoms described by patients, given that just one of only two specific disorders are possible. Reference 7 proposes a clustering method, which is somewhat different from the one outlined here, based on a type of algorithm called Isodata.

The class of Isodata algorithms considered all rely on defining a metric that supposedly defines a distance between subjects, which in this case are the patients. For

this purpose each patient is characterized as a vector with components determined by which symptoms the patient does or does not have.

The nonfuzzy versions of the algorithms are supposed to determine the membership of patients in the two nonfuzzy clusters, each of which consists of those patients with one of the two stomach disorders, by minimizing a quantity that depends monotonically on the distances between patients and a "centroid" vector associated with each cluster. The fuzzy version depends on maximizing a quantity that supposedly measures the quality of the partition of all patients into the two clusters.

Both types of algorithm involve lengthy computations because of their attempt to assign to the affinity among the subjects of a cluster an arbitrary metric that resembles the euclidean distance between points in a multidimensional space. This not only increases the amount of computation required, but it also runs the risk of biasing the result since no objective reason exists to justify the choice of such a metric.

The linchpin approach described earlier should work at least as well, even though its primary objective has been to discover the existence of natural clusters in data when the number of clusters is not predetermined. The stomach disorder diagnostic problem of Ref. 7 can be used to illustrate how the method might work for this type of application.

In the data given in Ref. 7 patients numbered 1 through 57 all have hiatal hernia, while the others, numbered 58 through 107, suffer from gallstones. In Tables 10 and 11, which list the patients in the two categories and the symptoms they have or have not experienced, for every patient number appearing in the first column, in each of 11 additional columns corresponding to 11 possible symptoms a 1 appears if the patient has the corresponding symptom and a 0 otherwise. The 11 symptoms are:

1. Male = 1; female = 0
2. Epigastric pain
3. Upper right quadrant pain
4. Back pain
5. Discomfort episodes of 1-4 weeks
6. Discomfort episodes of 0-1 days
7. Relief induced by food ingestion
8. Aggravation induced by food ingestion
9. Aggravation induced by position

10. Weight loss (at least 20 lb in 6 mos.)

11. Persistence of pain (at least 1 month in length).

Table 10. Symptoms of Patients With Hiatal Hernia

Patient	Symptoms										
1	0	1	0	0	0	1	0	0	0	0	0
2	0	1	0	0	0	1	0	0	0	0	0
3	0	1	0	0	0	1	0	1	0	0	0
4	0	1	0	0	0	1	0	1	0	0	0
5	0	1	0	0	0	1	0	1	0	0	0
6	0	1	0	0	0	1	0	1	0	0	0
7	0	1	0	0	0	1	0	1	1	0	0
8	0	1	0	0	0	1	0	1	1	0	0
9	0	1	0	0	0	1	0	1	1	0	0
10	0	1	0	0	0	1	1	0	0	0	0
11	0	1	0	0	0	1	1	0	1	0	0
12	0	1	0	0	0	1	1	0	1	0	1
13	0	1	0	0	0	1	1	1	0	0	0
14	0	1	0	0	1	1	0	1	0	0	1
15	0	1	0	1	0	1	0	0	1	0	0
16	0	1	0	1	0	1	0	0	1	0	1
17	0	1	0	1	0	1	0	1	0	1	1
18	0	1	0	1	0	1	0	1	1	0	0
19	0	1	0	1	0	1	1	0	1	0	0
20	0	1	0	1	0	1	1	0	1	0	0
21	0	1	0	1	0	1	1	0	1	0	0
22	0	1	0	1	1	0	0	1	1	0	1
23	0	1	1	0	0	0	0	1	1	0	1
24	0	1	1	0	1	0	0	0	0	0	0
25	0	1	1	0	1	0	1	0	0	1	1
26	0	1	1	1	1	0	0	1	0	0	0
27	1	0	0	0	0	0	0	0	0	0	0

(continued)

Table 10. Symptoms of Patients With Hiatal Hernia (continued)

Patient	Symptoms										
28	1	1	0	0	0	0	0	0	0	0	0
29	1	1	0	0	0	0	0	0	1	0	0
30	1	1	0	0	0	0	0	0	1	0	0
31	1	1	0	0	0	0	1	0	0	0	0
32	1	1	0	0	0	0	1	0	1	0	0
33	1	1	0	0	0	1	0	0	1	0	0
34	1	1	0	0	0	1	0	0	1	0	0
35	1	1	0	0	0	1	0	1	0	0	0
36	1	1	0	0	0	1	0	1	0	0	0
37	1	1	0	0	0	1	0	1	1	0	0
38	1	1	0	0	0	1	0	1	1	0	0
39	1	1	0	0	0	1	1	0	0	0	0
40	1	1	0	0	0	1	1	0	0	0	0
41	1	1	0	0	0	1	1	0	1	0	0
42	1	1	0	0	0	1	1	0	1	0	0
43	1	1	0	0	0	1	1	0	1	0	0
44	1	1	0	0	1	0	0	0	1	0	0
45	1	1	0	0	1	0	1	0	0	0	0
46	1	1	0	0	1	0	1	0	0	0	0
47	1	1	0	0	1	0	1	0	1	0	0
48	1	1	0	0	1	1	1	1	1	0	0
49	1	1	0	1	0	0	0	0	1	0	1
50	1	1	0	1	0	1	0	0	0	0	0
51	1	1	0	1	0	1	0	0	0	0	1
52	1	1	0	1	0	1	0	1	0	0	0
53	1	1	0	1	1	0	1	0	0	0	0
54	1	1	0	1	1	0	1	1	1	0	1
55	1	1	1	0	0	1	0	1	1	0	0
56	1	1	1	0	0	1	1	1	1	0	0
57	1	1	1	0	1	1	1	0	0	0	1

Table 11. Symptoms of Patients with Gallstones

Patient	Symptoms									
58	0	0	1	0	0	0	0	1	0	1	1
59	0	0	1	0	0	1	0	1	0	0	0
60	0	0	1	0	0	1	0	1	0	0	1
61	0	0	1	0	1	0	0	1	0	1	0
62	0	0	1	1	0	1	0	0	0	0	0
63	0	0	1	1	0	1	0	0	0	0	0
64	0	0	1	1	0	1	0	1	0	0	0
65	0	0	1	1	0	1	0	1	0	0	0
66	0	0	1	1	0	1	0	1	0	0	0
67	0	1	0	0	0	0	1	0	0	0	0
68	0	1	0	0	0	1	0	0	0	0	0
69	0	1	0	0	0	1	0	1	0	0	0
70	0	1	0	0	0	1	0	1	0	0	0
71	0	1	0	0	0	1	1	0	0	0	0
72	0	1	0	0	0	1	1	1	0	0	0
73	0	1	0	0	0	1	1	1	0	0	0
74	0	1	0	1	0	1	0	0	0	0	0
75	0	1	0	1	1	1	1	1	0	0	0
76	0	1	1	0	0	0	0	1	0	0	1
77	0	1	1	0	0	1	0	1	0	0	0
78	0	1	1	0	0	1	0	1	0	0	0
79	0	1	1	0	0	1	0	1	0	1	0
80	0	1	1	0	0	1	0	1	0	1	1
81	0	1	1	1	0	1	0	1	0	0	0
82	0	1	1	1	0	1	0	1	0	0	0
83	0	1	1	1	0	1	0	1	0	0	0
84	0	1	1	1	0	1	0	1	0	0	0
85	0	1	1	1	0	1	0	1	0	0	0
86	0	1	1	1	0	1	0	1	0	0	0

(continued)

Table 11. Symptoms of Patients with Gallstones (continued)

Patient	Symptoms										
87	0	1	1	1	0	1	0	1	0	0	1
88	0	1	1	1	0	1	0	1	0	1	0
89	1	1	1	1	0	1	1	1	0	0	0
90	1	0	1	0	0	1	0	1	0	0	0
91	1	0	1	1	0	1	0	1	0	0	0
92	1	0	1	1	0	1	0	1	0	0	0
93	1	0	1	1	0	1	0	1	0	0	0
94	1	1	0	0	0	1	0	0	0	0	0
95	1	1	0	0	0	1	0	0	1	0	0
96	1	1	0	0	0	1	0	1	0	0	0
97	1	1	0	0	0	1	0	1	1	0	0
98	1	1	0	0	0	1	0	1	1	1	0
99	1	1	0	0	1	1	0	1	1	0	0
100	1	1	0	1	0	1	0	0	0	0	0
101	1	1	1	0	0	1	0	0	0	0	1
102	1	1	1	0	0	1	0	1	0	0	0
103	1	1	1	0	0	1	0	1	0	0	0
104	1	1	1	0	0	1	0	1	0	0	1
105	1	1	1	0	0	1	1	0	0	0	0
106	1	1	1	1	0	1	0	0	0	0	0
107	1	1	1	1	0	1	0	0	0	0	0

An examination of Tables 10 and 11 reveals that there were 8 patients with hiatal hernia, each of whose symptoms were all identical to those that some patient with gallstones experienced. Therefore, it is immediately evident that the 11 symptoms are inadequate for distinguishing between the two possible disorders, at least for 16 of the 107 patients contributing to the data.

As a result, any diagnostic scheme based on the given symptoms must lead to some errors; the optimum algorithm for separating the patients into two clusters should then be regarded as the one that leads to the fewest errors.

An optimum algorithm based on the linchpin approach to fuzzy clustering is fairly obvious. Treating each patient as a possible linchpin, for the membership function of the associated cluster: (1) calculate the sum S_1 of the membership function values corresponding to those patients with hiatal hernia and the sum S_2 of the values corresponding to those patients with gallstones; (2) calculate $|S_1 - S_2|$; (3) choose as the optimum linchpin patient the one for which $|S_1 - S_2|$ has the largest value; (4) for the associated cluster the patients corresponding to the membership values that are above 0.5 will be diagnosed as having the same disorder as the optimal linchpin patient, and the rest as having the other disorder.

An application of the algorithm leads to the result that patient number 47 is an optimal linchpin. An examination of Table 10 shows that patient 47 has just the symptoms 1, 2, 5, 7, and 9. The membership function for the fuzzy cluster of patients exhibiting just these symptoms is given by:

0.54/1 + 0.54/2 + 0.45/3 + 0.45/4 + 0.45/5 + 0.45/6 + 0.54/7 + 0.54/8 + 0.54/9 + 0.64/10 + 0.73/11 + 0.64/12 + 0.54/13 + 0.45/14 + 0.54/15 + 0.45/16 + 0.18/17 + 0.45/18 + 0.64/19 + 0.64/20 + 0.64/21 + 0.54/22 + 0.45/23 + 0.64/24 + 0.54/25 + 0.45/26 + 0.64/27 + 0.73/28 + 0.82/29 + 0.82/30 + 0.82/31 + 0.91/32 + 0.73/33 + 0.73/34 + 0.54/35 + 0.54/36 + 0.64/37 + 0.64/38 + 0.73/39 + 0.73/40 + 0.82/41 + 0.82/42 + 0.82/43 + 0.91/44 + 0.91/45 + 0.91/46 + 1.00/47 + 0.82/48 + 0.64/49 + 0.54/50 + 0.45/51 + 0.45/52 + 0.82/53 + 0.73/54 + 0.54/55 + 0.64/56 + 0.64/57 + 0.18/58 + 0.27/59 + 0.18/60 + 0.36/61 + 0.27/62 + 0.27/63 + 0.18/64 + 0.18/65 + 0.18/66 + 0.73/67 + 0.54/68 + 0.45/69 + 0.45/70 + 0.64/71 + 0.54/72 + 0.54/73 + 0.45/74 + 0.54/75 + 0.36/76 + 0.36/77 + 0.36/78 + 0.27/79 + 0.18/80 + 0.27/81 + 0.27/82 + 0.27/83 + 0.27/84 + 0.27/85 + 0.27/86 + 0.18/87 + 0.18/88 + 0.36/89 + 0.36/90 + 0.27/91 + 0.27/92 + 0.27/93 + 0.64/94 + 0.73/95 + 0.54/96 + 0.54/97 + 0.54/98 + 0.73/99 + 0.54/100 + 0.45/101 + 0.45/102 + 0.45/103 + 0.36/104 + 0.64/105 + 0.45/106 + 0.45/107.

Keeping in mind that patients numbered less than 58 all have hiatal hernia and the rest all have gallstones, an examination of this membership function reveals that if a 50 percent or greater membership value is the requirement for deciding that a patient has hiatal hernia, 25 diagnostic errors will occur. The analysis of Ref. 7, which applied Isodata algorithms based on two different definitions of the norm for measuring distance between clusters, resulted in 23 errors for one and 25 errors for the other. Thus, for this

example no real difference exists between the effectiveness of the approach recommended in Ref. 7 and that of the algorithm suggested here.

Given the symptoms associated with the linchpin patient, it is not difficult to formulate an algorithm to determine which symptoms are the most useful for separating the patients into the hiatal hernia and gallstone clusters. For each symptom, the number of patients for whom the existence or nonexistence of the symptom is the same as it is for the linchpin patient is counted separately for patients numbered below 58 and for patients numbered above 57. The absolute value of the difference between the two counts can be regarded as a measure of the diagnostic value of the symptom.

The result of the algorithm applied with patient 47 as the linchpin is that symptom number 3, upper right quadrant pain, suffered by a patient with gallstones but not by one with hiatal hernia, is the clear winner. It is followed by symptom 9, aggravation induced by position, suffered by a patient with hiatal hernia but not one with gallstones. A close third is number 8, aggravation induced by food ingestion, suffered by a patient with gallstones but not by one with hiatal hernia. Reference 7 arrives at the identical conclusions using a similar algorithm applied to the centroid vectors calculated by the Isodata algorithms.

An inspection of Tables 10 and 11 reveals immediately that using symptom 3 alone as the diagnostic criterion causes misdiagnoses for 7 hiatal hernia and 16 gallstone patients: a total of 23. Using 3 with either 8 or 9, or with both, as the criterion causes 36 misdiagnoses when the Isodata algorithm is used and 27 with the linchpin approach. These results indicate that symptom 3 alone is an optimal diagnostic criterion: using more symptoms does not reduce, and may increase, the number of misdiagnoses.

REFERENCES

1. L.A. Zadeh, "Fuzzy Sets," *Informat. Control*, V. 8, 1965, pp. 338-353.

2. R. Sutton and D.R. Towill, "An Introduction to the Use of Fuzzy Sets in the Implementation of Control Algorithms," *J. Inst. Electronic and Radio Engs.*, V. 55, No. 10, 1985, pp. 357-367.

3. S.Yasunobu and S. Miyamoto, "Automatic Train Operation System by Predictive Fuzzy Control," *Industrial Applications of Fuzzy Control* (M. Sugeno, Ed.), Elsevier Science Publishers B.V. (North-Holland), 1985, pp. 1-18.

4. R.F. Popoli and J.M. Mendel, "Heuristically Constrained Estimation for Intelligent Signal Processing," *Advances in Geophysical Data Processing--Artificial Intelligence and Expert Systems in Petroleum Exploration* (Eds. M. Simaan and F. Aminzadeh), V. 3, Jai Press Inc., London, 1989, pp. 107-134.

5. G.W. Deley, *Radar Handbook* (M.I. Skolnik, Ed.), Ch. 3, McGraw-Hill, New York, 1970, p. 3-8 ff.

6. M. Sugeno (Ed.), *Industrial Applications of Fuzzy Control,* Elsevier Science Publishers B.V. (North-Holland), 1985.

7. J.C. Bezdek, *Feature Selection for Binary Data--Medical Diagnosis with Fuzzy Sets*, National Computer Conference, 1976, pp. 1057-1068.

8. J.A. Hartigan, *Clustering Algorithms*, Chs. 3 and 9, John Wiley & Sons, New York, 1975.

APPENDIX A

FUZZY KALMAN FILTER

APPENDIX A
FUZZY KALMAN FILTER

I. INTRODUCTION

This appendix outlines a possible application of fuzzy sets to the design of a Kalman filter for use in tracking a maneuvering airborne target by means of a ground based radar. For simplicity the tracking problem discussed here is two-dimensional: i.e., the target trajectory remains in a plane orthogonal to the earth's surface and containing the radar.

The assumed scenario is that the target first appears at the horizon in some direction. It moves toward the radar, passing it overhead, and continues on toward the horizon in the opposite direction. Otherwise the scenario assumes that no a priori knowledge of the target trajectory is available.

The design of a Kalman filter ordinarily depends on having an equation of motion, called the system model, to predict future values of the so-called state vector that defines the target behavior. In the present case no such equation exists in the usual sense.

However, certain geometrical relationships that are independent of the target's detailed trajectory do exist, and it should be possible to formulate a kind of system model in terms of them. These relationships are necessarily fuzzy in nature; therefore, any system model derived from them will also be fuzzy.

This document indicates how such a system model might be characterized. It also suggests an algorithm for implementing the corresponding fuzzy Kalman filter.

The next section offers a quick review of the standard Kalman filter theory, as presented, for example, in Ref. 1. Section III defines an algorithm that may be used to implement the fuzzy Kalman filter suggested for the tracking problem considered here.

II. KALMAN FILTERS

A Kalman filter is a recursive process that performs repeated least square error estimates of a set of time varying parameters, using the following steps:

1. Based on a theoretical model for the behavior of the parameters over time, it predicts new parameter values, as well as an error covariance matrix, a short time interval ahead;

2. Using the error covariance matrix and measurements furnished by a sensor to verify the prediction, it updates the estimated parameter values;

3. With the same information it calculates an updated estimate of the error covariance matrix;

4. It repeats the preceding steps.

The relation between the measurements referred to in step 2 and the parameter values in the presence of noise, which is assumed to be additive, is called the measurement model. If the theoretical model of the parameter behavior, also called the system model, and the measurement model are both linear, they take the form of the following equations in a discrete version of the Kalman filter:

System Model-- $$\underline{x}_k = \Phi_{k-1}\underline{x}_{k-1} + \underline{w}_{k-1}, \tag{1}$$

where the subscripts indicate measurement times, the underlined quantities are vectors, the Φ_{k-1} are matrices, and the $\underline{w}_k$ are noise vectors for which the components are assumed to be independent, Gaussian distributed random variables with mean zero and variances equal to the elements in the main diagonal of a diagonal matrix Q_k;

Measurement Model-- $$\underline{z}_k = H_k\underline{x}_k + \underline{v}_k, \tag{2}$$

where the underlined quantities are vectors, the H_k are matrices, and the ν_k are noise vectors for which the components are independent, Gaussian distributed random variables with mean zero and variances equal to the elements in the main diagonal of a diagonal matrix R_k.

The expected values of the so-called state vectors $\underline{x}_k$ are expressed by

$$E[\underline{x}_k] = \hat{\underline{x}}_k. \tag{3}$$

In this notation the error covariance is defined by

$$E[(\hat{\underline{x}}_k - \underline{x}_k)(\hat{\underline{x}}_k - \underline{x}_k)^T] = P_k, \tag{4}$$

where the superscript T means the transpose of a matrix, which in this case is a vector regarded as a 1 by n matrix.

The mean value of equation (1) is used to extrapolate the expected value of a state vector from an updated estimate of the vector after a measurement; i.e.,

$$\hat{x}_k(-) = \Phi_{k-1}\hat{x}_{k-1}(+), \tag{5}$$

where (+) indicates an expected state vector that has been updated after a measurement and (–) indicates the expected state vector extrapolated to the future next measurement time. A similar extrapolation of the error covariance matrix is given by the equation

$$P_k(-) = \Phi_{k-1}P_{k-1}(+)\Phi_{k-1}^T+Q_{k-1}. \tag{6}$$

The updated expected state vector, after incorporating the information obtained from a measurement, is given by

$$\hat{x}_k(+) = \hat{x}_k(-)+K_k[z_k-H_k\hat{x}_k(-)], \tag{7}$$

where K_k is the so-called Kalman gain matrix, given by

$$K_k = P_k(-)H_k^T[H_kP_k(-)H_k^T+R_k]^{-1}. \tag{8}$$

Finally, the error covariance matrix updated on the basis of the measurement information is given by

$$P_k(+) = [I-K_kH_k]P_k(-), \tag{9}$$

where I is the identity matrix.

Since the system model never predicts exact behavior, which in practice may deviate considerably from the prediction, the Kalman filter performance in a real environment is usually degraded from what is theoretically anticipated. The associated discrepancy is called divergence.

In some cases the filter is unstable, so that the divergence tends to grow without limit. A common method of avoiding this is to add an artificial noise term to the system model, a procedure that has the effect of causing the filter to rely more heavily on the measurement data, thereby reducing the effective weight given to the system model extrapolation. If the added noise is sufficiently large the divergence will at least remain bounded, even if the measurement error is larger than what might be desired.

III. A FUZZY KALMAN FILTER

An example for which a reliable system model may be impossible to formulate is the problem of tracking a maneuvering target, since to some extent the target's motion will be arbitrary. Adding a noise term to a system model prediction, which would otherwise amount to an assumed target trajectory, is equivalent to guessing that deviations from the trajectory can be regarded as noise-like, i.e., *random in nature.*

Since the added noise term is *ad hoc* in any case and the deviation it is supposed to represent may not even be stochastic, a better solution might be to replace the system model with one that is fuzzy, i.e., a model expressed in terms of fuzzy algorithm relational statements. In this case (1) would be replaced by statements of the form "If $\underline{x}_k$ is $\underline{A}_k$ then $\underline{x}_{k+1}$ is $\underline{B}_k$," which can be expressed symbolically as

$$(\underline{x}_k \in \underline{A}_k) \rightarrow (\underline{x}_{k+1} \in \underline{B}_k). \tag{10}$$

The symbols $\underline{A}_k$ and $\underline{B}_k$ designate fuzzy sets, which are characterized by membership functions $\mu_{\underline{A},k}(u)$ and $\mu_{\underline{B},k}(v)$ over appropriate universes of discourse. To implement the fuzzy model as a replacement of (5) for the purpose of extrapolating the expected state vector, it is necessary to form the fuzzy relation matrix $\mathbf{R}_k$ corresponding to rule expressed in (10):[1]

$$\mathbf{R}_k = \begin{bmatrix} \min[\mu_{A,k}(u_1), \mu_{B,k}(v_1)] & \cdots & \min[\mu_{A,k}(u_1), \mu_{B,k}(v_n)] \\ \vdots & & \vdots \\ \min[\mu_{A,k}(u_m), \mu_{B,k}(v_1)] & \cdots & \min[\mu_{A,k}(u_m), \mu_{B,k}(v_n)] \end{bmatrix}, \tag{11}$$

wherein it is assumed that the universe of discourse for the set A has m possible values and that for the set B has n possible values. In carrying out the min operation in (11) to determine the elements of the matrix $\mathbf{R}_k$, it should be kept in mind that the quantities involved are vectors: min[$\underline{A}$,$\underline{B}$] means the vector for which each component is defined to be the smaller of the two corresponding components of $\underline{A}$ and $\underline{B}$.

The fuzzy extrapolation replacing (5) assumes that a fuzzy updated expected state vector at the time k has been estimated. Then the membership function corresponding to the expected state extrapolated to the time k+1 can be calculated by means of the composition rule:

$$\mu_{\underline{x},k+1}(v) = \max_u \min[\mu_{\underline{x},k}(u), \mu_{\underline{R}}(u,v)], \tag{12}$$

where $\mu_{\underline{R}}(u,v)$ represents the elements of the relational matrix R defined in (11).

To replace the fuzzy state vector calculated by means of (12) with a crisp (non-fuzzy) estimate, the recommended approach is to use the centroid of the fuzzy vector membership function relative to the universe of discourse. Then the lth component of the crisp state vector would be given by

---

[1] Cf. Ref. 2.

$$\hat{x}_{kl} = \frac{\sum_{v} v\,\mu_{\hat{x},kl}(v)}{\sum_{v} \mu_{\hat{x},kl}(v)}. \tag{13}$$

In the discussion of the fuzzy Kalman filter, for simplicity it will be assumed that the matrix H_k in the measurement model is the identity; i.e., instead of (2)

$$z = \underline{x}_k + \underline{v}_k. \tag{14}$$

It should not be too difficult, however, to extend the results to arbitrary H_k if the need should arise. It will also be assumed that measurements are all crisp, i.e., have distinct numerical values.[2]

In the case of a fuzzy Kalman filter a fuzzy system model based on (10) will replace the crisp system model based on (1). The fuzzy extrapolated state vector, which replaces the extrapolated crisp state vector given by (5), will be determined by (10), (11), (12), and (13).

Note that it is not necessary to calculate an expected value as in (3) because in the fuzzy case no stochastic process noise is assumed. For this reason a caret over any quantity $\underline{x}$ with a membership function $\mu_x(v)$ is defined to be an operator that transforms $\underline{x}$ into an "extrapolation," of which the l^{th} component is given by

$$\hat{x}_l = \frac{\sum_{v} v\,\mu_{\hat{x},l}(v)}{\sum_{v} \mu_{\hat{x},l}(v)}, \tag{15}$$

where

$$\mu_{\hat{x}}(v) = \max_{u} \min\left[\mu_x(u), \mu_R(u,v)\right] \tag{16}$$

and $\mu_{\hat{x},l}(v)$ is the l^{th} component of $\mu_{\hat{x}}$.

If $\underline{x}$ is a crisp quantity its membership function is 0 at every value in the universe of discourse except at the one equal to $\underline{x}$, itself, where the membership function is 1. Then the extrapolated value of $\underline{x}$ will be given by (15) with

$$\mu_{\hat{x}}(v) = \min\left[\mu_{A,k}(u_j), \mu_{B,k}(v)\right], \tag{17}$$

---

2 If the measurements were characterized as fuzzy sets no stochastic error terms would remain, and the minimum error variance optimization criterion imposed in deriving the Kalman filter algorithm would not be meaningful.

where u_j is such that

$$\mu_x(u_j) = 1;$$

i.e., the l^{th} component of the extrapolated vector will be given by

$$\hat{x}_l = \frac{\sum_v v \min[\mu_{A,k}(u_j), \mu_{B,k}(v)]_l}{\sum_v \min[\mu_{A,k}(u_j), \mu_{B,k}(v)]_l}. \tag{18}$$

In the fuzzy Kalman filter algorithm, the next step after replacing (5) with a fuzzy extrapolation of the state vector is to substitute for (6) a fuzzy rule for extrapolating the error covariance. This can be done by referring to the definition (4), which implies that

$$P_{k-1}(+) = E[(\hat{x}_{k-1}(+) - x_{k-1})(\hat{x}_{k-1}(+) - x_{k-1})^T], \tag{19}$$

where $\hat{\underline{x}}_{k-1}$ is the result of a fuzzy extrapolation based on (10), (11), (12), and (13). On the assumption that the measurement errors are small compared to the uncertainties in the fuzzy predictions, the measured value of the state vector can be substituted for its actual value, so that (19) becomes

$$P_{k-1}(+) = (\hat{x}_{k-1}(+) - z_{k-1})(\hat{x}_{k-1}(+) - z_{k-1})^T. \tag{20}$$

Taking into account that no process noise Q_{k-1} exists, a comparison of (20) with the right hand side of (6) indicates that the appropriate generalization of (6) will be given by

$$P_k(-) = F\{\hat{x}_{k-1}(+) - z_{k-1}\} F\{\hat{x}_{k-1}(+) - z_{k-1}\}^T, \tag{21}$$

where $F\{\underline{V}_{k-1}\}$ means the fuzzy extrapolation of a vector quantity $\underline{V}_{k-1}$ by means of (10), (11), (12), and (13). Since the quantities being extrapolated on the right hand sides of (20) and (21) are all crisp, (13) will take the special case form determined by (17) and (18).

The rule given by (8) determines the Kalman gain matrix just as it does in the non-fuzzy case. Since it is assumed here that the matrix H_k is the identity the rule specializes to

$$K_k = P_k(-)[P_k(-) + R_k]^{-1}. \tag{22}$$

The state vector updating rule (7) is also the same for the fuzzy as it is for the non-fuzzy case, specializing to

$$\hat{x}_k(+) = \hat{x}_k(-) + K_k[z_k - \hat{x}_k(-)], \tag{23}$$

as is the error covariance matrix updating rule (9), which specializes to

$$P_k(+) = [I - K_k]P_k(-). \tag{24}$$

An example to illustrate how a fuzzy Kalman filter might work in practice is the problem of a ground based radar attempting to track a maneuvering airborne target. If the target's maneuvers are controlled by a human they will neither be predictable nor random.

If enough deviations from a predictable constant altitude flight path exist, a fuzzy system model should be more realistic than an artificial one (e.g., based on assuming constant altitude and speed) augmented by process noise. In fact, one would expect the type of fuzzy Kalman filter described here to provide an appropriate tracking algorithm.

To keep the example simple it will be assumed that the target remains in a vertical plane containing the radar. Then the range RT and elevation angle ψ determine the target position at any time.

Assume that the target flight begins beyond the horizon relative to the radar. Then despite the lack of prior information about its trajectory, certain facts concerning the behavior of ψ and R_T, nevertheless, do exist:

(1) if the target elevation angle is near 0° or 180° then it increases (or decreases) at a slow rate and the range is large;

(2) if the elevation angle is near 90° then it increases (or decreases) at a rapid rate and the range is small.

More formally,

$$\begin{aligned}&(1)\quad [(\psi \sim 0^\circ) \vee (\psi \sim 180^\circ)] \rightarrow (R_T \text{ large}) \rightarrow \left(\left|\frac{d\psi}{dt}\right| \text{ small}\right);\\&(2)\quad (\psi \sim 90^\circ) \rightarrow (R_T \text{ small}) \rightarrow \left(\left|\frac{d\psi}{dt}\right| \text{ large}\right).\end{aligned} \tag{25}$$

On the other hand, for quantities involved in a discrete recursive relation defining a fuzzy system model analogous to the recursive relation indicated in equation (1),

$$\left(\left|\frac{d\psi}{dt}\right| \text{ small}\right) \rightarrow \left[\operatorname{sgn}(\psi_{k+1} - \psi_k) = \operatorname{sgn}\left(\frac{d\psi}{dt}\right)\right] \wedge \begin{cases}(\psi_k \sim 0^\circ) \rightarrow (\psi_{k+1} \sim 0^\circ),\\(\psi_k \sim 180^\circ) \rightarrow (\psi_{k+1} \sim 180^\circ),\\(\psi_k \sim 90^\circ) \rightarrow (\psi_{k+1} \sim 90^\circ)\end{cases} \tag{26}$$

and

$$\left(\left|\frac{d\psi}{dt}\right| \text{ large}\right) \rightarrow \left[\operatorname{sgn}(\psi_{k+1} - \psi_k) = \operatorname{sgn}\left(\frac{d\psi}{dt}\right)\right] \wedge [(\psi_k < \sim 90^\circ) \rightarrow (\psi_{k+1} > \sim 90^\circ)]. \tag{27}$$

Finally, using (26) and (27) in (25)

$$
\begin{aligned}
&(1)\ [(\psi_k = 0°) \vee (\psi_k = 180°)] \rightarrow (R_{Tk} \text{ large}),\\
&(2)\ (\psi_k = 90°) \rightarrow (R_{Tk} \text{ small}),\\
&(3)\ (\psi_k = 0°) \rightarrow (\psi_{k+1} = 0°),\\
&(4)\ (\psi_k = 180°) \rightarrow (\psi_{k+1} = 180°),\\
&(5)\ (\psi_k < -90°) \rightarrow (\psi_{k+1} > -90°),\\
&(6)\ \sigma_k < 0 \rightarrow \psi_{k+1} < \psi_k,\\
&(7)\ \sigma_k > 0 \rightarrow \psi_{k+1} > \psi_k,
\end{aligned}
\tag{28}
$$

where

$$\sigma_k = \operatorname{sgn}\left(\frac{d\psi}{dt}\right)_k .$$

The state vector $\underline{x}_k$ has the components ψ_k, σ_k, and R_{Tk}. The seven fuzzy relations in (28) determine the fuzzy system model, except for the extrapolation of σ_k to σ_{k+1}. Since most of the time

$$\sigma_k \rightarrow (\sigma_{k+1} = \sigma_k) \tag{29}$$

will be true, (29) will be included among the rules defining the system model.

The first step in constructing the desired fuzzy system model is to form the relation matrix, defined by (11), corresponding to each of the rules in (28) for extrapolating from ψ_k to ψ_{k+1}. Because those are the rules labeled (3), (4), and (5), the associated relation matrices will be designated by $\mathbf{R}_{k,3}$, $\mathbf{R}_{k,4}$, and $\mathbf{R}_{k,5}$. The fuzzy logical sum of corresponding elements in those three matrices then forms each element of the combined relation matrix $\mathbf{R}_{k,\psi}$ for extrapolating the state vector component ψ_k. That is, each element of $\mathbf{R}_{k,\psi}$ is the largest of the three corresponding elements of $\mathbf{R}_{k,3}$, $\mathbf{R}_{k,4}$, and $\mathbf{R}_{k,5}$. In summary:

$$\mathbf{R}_{k,\psi} = \mathbf{R}_{k,3} \vee \mathbf{R}_{k,4} \vee \mathbf{R}_{k,5}, \tag{30}$$

where each element $R_{k,\psi,ij}$ of $\mathbf{R}_{k,\psi}$ is given by

$$R_{k,\psi,ij} = \max[R_{k,3,ij},\ R_{k,4,ij},\ R_{k,5,ij}] \tag{31}$$

and each element of the $\mathbf{R}_{k,\psi}$ on the right hand side of (31) is given by

$$R_{k,v,ij} = \min[\mu_{v,k}(u_i)\ ,\ \mu_{v,k+1}(v_j)] \tag{32}$$

in accordance with (11).

In a similar manner, rules labeled (1) and (2) in (28) determine a relation matrix for extrapolating from ψ_{k+1} to R_{Tk+1} after extrapolating from ψ_k to ψ_{k+1}. It should be remembered that $\vee$ (the logical *or*) in rule (1) implies that the fuzzy set membership

functions associated with the .wo approximate values, 0° and 180°, of ψ_k must be combined into a single membership function formed by choosing for it the larger of the two membership function values at every element in the universe of discourse.

Rules (6) and (7) are used, in place of (7), for updating the component s_k after a measurement. The rule for extrapolating σ_k to σ_{k+1} between measurements is (29).

Each relation set off by parentheses in (28) implies a fuzzy set for which a satisfactory membership function can be defined simply. For example the set given by

$$\psi < \sim 90°$$

might have a membership function depicted by the curve in Figure A-1. The set given by

$$\psi > \sim 90°$$

might then have the membership function depicted by the curve in Figure A-2.

IV. CONCLUSIONS

In the absence of a specific system model, general physical or geometrical considerations based on experience and common sense may be used to formulate a fuzzy system model for use in designing a Kalman filter based on the concept of fuzzy sets. The construction of the associated algorithm should then be possible by means of standard procedures already developed for other applications of fuzzy set theory.

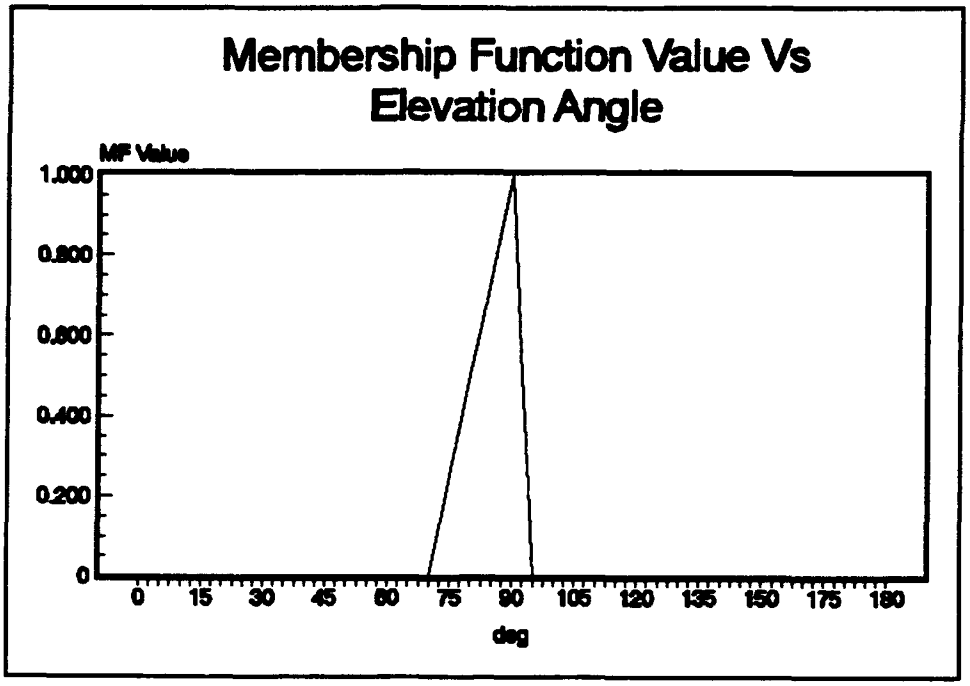

Figure A-1. Membership Function for ψ <~90°

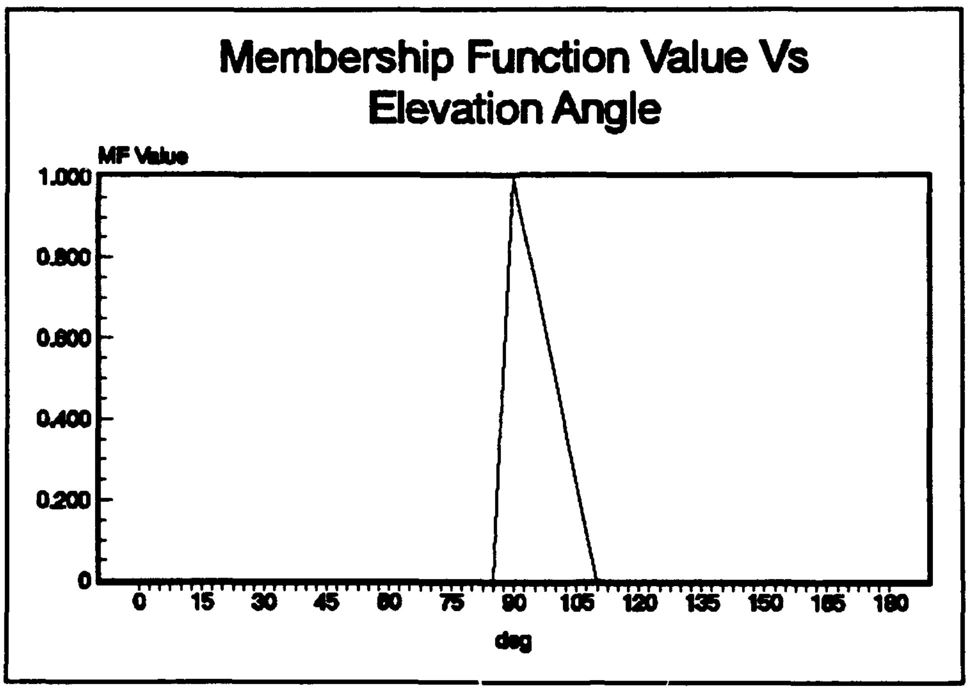

Figure A-2. Membership Function for ψ >~90°

REFERENCES

A-1. A. Gelb (Ed.) et al., "Applied Optimal Estimation," M.I.T. Press, Cambridge (1974).

A-2. R. Sutton and D.R. Towill, "An Introduction to the Use of Fuzzy Sets in the Implementation of Control Algorithms," *J. Inst. Elect. and Radio Eng.*, Vol. 55, No. 10, 1985, pp. 357-367.